20 March, 2009 (Version 6) 2nd circulation
Patric.Amcoff@oecd.org

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

DRAFT PROPOSAL FOR A NEW GUIDELINE

In Vitro Skin Irritation: Reconstructed Human *Epidermis* (RhE) Test Method

INTRODUCTION

1. Skin irritation refers to the production of reversible damage to the skin following the application of a test chemical for up to 4 hours [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS)](1). This Test Guideline provides an *in vitro* procedure that, depending on country/regional regulatory requirements, may allow determining the skin irritancy of chemicals as a replacement test to the rabbit *in vivo* test (2), as a screening test, or within a tiered testing strategy in a weight of evidence approach.

2. The assessment of skin irritation has typically involved the use of laboratory animals (OECD Test Guideline 404; adopted in 1981 and revised in 1992 and 2002)(2). In relation to animal welfare concerns, TG 404 was revised in 2002, allowing for the determination of skin corrosion/irritation by applying a tiered testing strategy, using validated *in vitro* or *ex vivo* test methods, thus avoiding pain and suffering of animals[1].

3. This Test Guideline is based on a reconstructed human *epidermis* (RhE) model, which in its overall design (the use of human derived *epidermis* keratinocytes as cell source and use of representative tissue and cytoarchitecture) closely mimics the biochemical and physiological properties of the upper parts of the human skin, *i.e.,* the *epidermis*. The test method described under this Test Guideline allows the hazard identification of irritant chemicals in accordance with GHS category 2, as well as the identification of GHS no category chemicals, depending on country/regional regulatory requirements (for member states that do not adopt optional category 3)(1). This Test Guideline also includes a set of Performance Standards (PS)(Annex 2) for the assessment of similar and modified RhE-based test methods (6), in accordance with the principles of Guidance Document No. 34 (7).

4. Prevalidation, optimisation and validation studies have been completed for an *in vitro* test method (8)(9)(10)(11)(12)(13)(14)(15)(16)(17), commercially available as EpiSkin™ using an RhE model. Based on the acknowledged validity of EpiSkin™ (designated the Validated Reference Method - VRM), this Test Guideline is recommended as a replacement test method for the rabbit *in vivo* test (2), as a screening test, or as part of a tiered testing strategy in a weight of evidence approach, for classifying GHS category 2 irritant chemicals, as well as GHS no category chemicals depending on country/regional regulatory requirements (for member states that do not adopt optional category 3).

5. Before a proposed similar or modified *in vitro* RhE test method for skin irritation can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure that it is similar to that of the VRM, in accordance with the requirements of the PS set out in this Test Guideline (Annex 2). A similar or modified test method should at least demonstrate a sensitivity of 80% and a specificity of 70% when being tested using the 20 recommended Reference Chemicals of the PS.

[1] Three validated *in vitro* test methods have been adopted as OECD Test Guidelines 430, 431 and 435 (3)(4)(5), to be used for the corrosivity part of the tiered testing strategy of TG 404.

6. Two other validated *in vitro* skin irritation RhE test methods have shown similar results to the VRM (18). These are the modified EpiDerm™ test method and the SkinEthic RHE™ test method (18). Based on their acknowledged validity, these test methods are equally recommended as replacement test methods for the rabbit *in vivo* test, as screening tests, or as part of a tiered testing strategy in a weight of evidence approach, for classifying GHS category 2 irritant chemicals, as well as GHS no category chemicals depending on country/regional regulatory requirements (for member states that do not adopt optional category 3).

INITIAL CONSIDERATIONS AND LIMITATIONS

7. A limitation of the Test Guideline is that it only classifies compounds as skin irritants according to GHS category 2 and it does not allow the classification of chemicals to the optional category 3, mild irritants (1). However, depending on country or regional regulatory requirements, all non-category 2 chemicals may be considered non-classified (no category). Thus, regulatory requirements in member countries will decide if this Test Guideline will be used as a replacement test, as a screening test, or as part of a tiered testing strategy in a weight of evidence approach. Depending on regulatory considerations in member countries, follow-up *in vivo* testing may be required to fully characterize skin irritation potential. It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

8. This Test Guideline allows the hazard identification of irritant mono-, and multi-constituent chemicals (19), but it does not provide adequate information on skin corrosion. Since chemicals representing a wide spectrum of chemical classes were included in the validation study of the VRM, the Test Guideline is expected to be generally applicable across chemical classes (16). The Test Guideline is applicable to solids, liquids, semi-solids and waxes. The liquids may be aqueous or non aqueous; solids may be soluble or insoluble in water. Solids should be ground to a fine powder before application; no other prior treatment of the sample is required. Gases, aerosols, formulations and preparations have not been assessed yet in a validation study. While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow their testing. It should also be noted that highly coloured chemicals, *e.g.,* hair dye components, may interfere with the cell viability measurements and need the use of adapted controls for corrections (see paragraphs 23-25).

DEFINITIONS

9. Definitions used are provided in Annex 1.

PRINCIPLE OF THE TEST

10. The test chemical is applied topically to a three-dimensional RhE model, comprised of normal, human-derived epidermal keratinocytes, which have been cultured to form a multilayered, highly differentiated model of the human *epidermis*. It consists of organized basal, spinous and granular layers, and a multilayered *stratum corneum* containing intercellular lamellar lipid layers arranged in patterns analogous to those found *in vivo*.

11. The principle of the RhE test method is based on the premise that chemicals are able to penetrate the *stratum corneum* and irritant chemicals are cytotoxic to the cells in the underlying layers. Cell viability is measured by dehydrogenase conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue; CAS RN 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (20). Irritant chemicals are identified by their ability

to decrease cell viability below defined threshold levels (*i.e.*, ≤ 50%, for GHS category 2). Depending on country/regional regulatory requirements and applicability of the Test Guideline, chemicals that produce cell viabilities above the defined threshold level, may be considered non-irritants (*i.e.,* > 50%, no category).

DEMONSTRATION OF PROFICIENCY

12. Prior to routine use of any of the three validated test methods (EpiSkin™ (VRM), EpiDerm™ and SkinEthic RHE™) that adhere to this Test Guideline, laboratories should demonstrate technical proficiency, using the ten recommended Proficiency Chemicals in Table 1. For similar (me-too) tests developed under this Test Guideline that are structurally and functionally similar to the VRM, or for modifications of any of the three validated test methods, the PS described in Annex 2 of this Test Guideline should be used to demonstrate similar reliability and accuracy of the test method prior to its use for regulatory testing.

13. As part of the proficiency exercise, it is recommended to verify the barrier properties of the tissues as specified by the test method producer. Once a test method has been successfully established and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties at regular intervals, *e.g.* every six months.

Table 1. Proficiency Chemicals[1]

Chemical	CAS RR	*In vivo* score[2]	Physical state	GHS category
naphthalene acetic acid	86-87-3	0	Solid	No Cat.
isopropanol	67-63-0	0.3	Liquid	No Cat.
methyl stearate	112-61-8	1	Solid	No Cat.
heptyl butyrate	5870-93-9	1.7	Liquid	No Cat. (*Optional Cat. 3*)[3]
hexyl salicylate	6259-76-3	2	Liquid	No Cat. (*Optional Cat. 3*)[3]
cyclamen aldehyde	103-95-7	2.3	Liquid	Cat. 2
1-bromohexane	111-25-1	2.7	Liquid	Cat. 2
potassium hydroxide (5% aq.)	1310-58-3	3	Liquid	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	3.3	Solid	Cat. 2
heptanal	111-71-7	4	Liquid	Cat. 2

[1] The Proficiency Chemicals are a subset of the chemicals used in the validation study.

[2] *In vivo* score in accordance with the OECD Test Guideline 404 (2).

[3] Under this Test Guideline, the GHS optional category 3 (1) is considered as no category.

PROCEDURE

14. The following is a description of the components and procedures of an RhE test method for skin irritation assessment. An RhE model should be reconstructed, and can be in-house-prepared or obtained

commercially (*e.g.,* EpiSkin™ (VRM), EpiDerm™ and SkinEthic RHE™). Standard Operating Procedures (SOPs) for the three validated test methods that adhere to this Test Guideline are available (21)(22)(23). Testing should be performed according to the following:

RHE TEST METHOD COMPONENTS

General conditions

15. Normal human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, *stratum spinosum, stratum granulosum*) should be present under a functional *stratum corneum*. *Stratum corneum* should be multilayered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic marker chemicals, *e.g.,* sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET_{50}) upon application of the marker chemical at a specified, fixed concentration. The containment properties of the RhE model should prevent the passage of material around the *stratum corneum* to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional conditions

Viability

16. The preferred assay for determining the magnitude of viability is the MTT (20). The optical density (OD) of the extracted (solubilised) dye from the tissue treated with the negative control (NC) should be at least 20 fold greater than the OD of the extraction solvent alone. It should be documented that the tissue treated with NC is stable in culture (provide similar viability measurements) for the duration of the test exposure period.

Barrier function

17. The *stratum corneum* and its lipid composition should be sufficient to resist the rapid penetration of cytotoxic marker chemicals, *e.g.,* SDS or Triton X-100, as estimated by IC_{50} or ET_{50}.

Morphology

18. Histological examination of the RhE model should be performed demonstrating human *epidermis*-like structure (including multilayered *stratum corneum*).

Reproducibility

19. The results of the test method should demonstrate reproducibility over time, preferably by an appropriate batch control (benchmark) chemical (see Annex 1).

Quality control (QC)

20. The RhE model developer/supplier should ensure that each batch of the RhE model used meets defined production release criteria, among which those for *viability* (paragraph 16), *barrier function* (paragraph 17) and *morphology* (paragraph 18) are the most relevant. These data should be provided to the

test method users, so that they are able to include this information in the test report. An acceptability range (upper and lower limit) for the IC_{50} or the ET_{50} should be established by the RhE model developer/supplier (or investigator when using an in-house model). Only results produced with qualified tissues can be accepted for reliable prediction of irritation classification. As an example, the acceptability ranges for the three validated test methods that adhere to this Test Guideline are given in Table 2.

Table 2. Examples of QC batch release criteria

	Lower acceptance limit	Mean of acceptance range	Upper acceptance limit
VRM (18 hours treatment with SDS)(21)	IC_{50} = 1.0 mg/ml	IC_{50} = 2.3 mg/ml	IC_{50} = 3.0 mg/ml
EpiDerm™ (1% Triton X-100)(22)	ET_{50} = 4.8 hr	ET_{50} = 6.7 hr	ET_{50} = 8.7 hr
SkinEthic RHE™ (1% Triton X-100)(23)	ET_{50} = 4.0 hr	ET_{50} = 5.6 hr	ET_{50} = 9.0 hr

Application of the Test and Control Chemicals

21. A sufficient number of tissue replicates should be used for each trial and for the controls (at least three replicates per run). For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the *epidermis* surface while avoiding an infinite dose, *i.e.,* a minimum of 25 $\mu L/cm^2$ or 25 mg/cm^2 should be used. For solid chemicals, the *epidermis* surface should be moistened with deionised or distilled water before application, to improve contact between the test chemical and the *epidermis* surface. Whenever possible, solids should be tested as a fine powder. At the end of the exposure period, the test chemical should be carefully washed from the *epidermis* surface with aqueous buffer, or 0.9% NaCl. Depending on which of the three validated RhE test methods that adhere to this Test Guideline is used, the exposure period varies between 15, 42 and 60 minutes, and the incubation temperature between 20 and 37°C. These exposure periods and temperatures are optimized for each RhE test method and represent the different intrinsic properties of the test methods, for further details, see the SOPs (21)(22)(23).

22. Concurrent NC and positive controls (PC) should be used for each study to demonstrate that viability (NC), barrier function and resulting tissue sensitivity (PC) of the tissues are within a defined historical acceptance range. The suggested PC chemical is 5% aqueous SDS. The suggested NC chemicals are water or phosphate buffered saline.

Cell Viability Measurements

23. The most important element of the test procedure is that viability measurements are not performed immediately after the exposure to the test chemicals, but after a sufficiently long post-treatment incubation period of the rinsed tissues in fresh medium. This period allows both for recovery from weak cytotoxic effects and for appearance of clear cytotoxic effects. The test method optimisation phase (9)(10)(11)(12)(13) demonstrated that a 42 hours post-treatment incubation period was optimal.

24. The MTT assay is a validated quantitative method which should be used to measure cell viability. It is compatible with use in a three-dimensional tissue construct. The tissue sample is placed in MTT solution of appropriate concentration (*e.g.,* 0.3 – 1 mg/mL) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (*e.g.,* isopropanol or acidic isopropanol), and the concentration of formazan is measured by determining the OD at 570 nm using a filter bandpass of maximum ± 30 nm.

25. Optical properties of the test chemical or its chemical action on the MTT may interfere with the assay leading to a false estimate of viability (because the test chemical may prevent or reverse the colour generation as well as cause it). This may occur when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the *epidermis*. If the test chemical acts directly on the MTT, is naturally coloured, or becomes coloured during tissue treatment, additional controls should be used to detect and correct for test chemical interference with the viability measurement technique. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the three validated test methods (21)(22)(23). Non-specific colour (NSC) due to these interferences should not exceed 30% of NC (for corrections). If NSC > 30%, the test chemical is considered as incompatible with the test method.

ACCEPTABILITY CRITERIA

26. For each test method using valid RhE model batches (see paragraph 22), tissues treated with the NC should exhibit OD reflecting the quality of the tissues that followed shipment, receipt steps and all protocol processes. Control OD values should not be below historical established boundaries. Similarly, tissues treated with the PC, should reflect their ability to respond to an irritant chemical in the conditions of the test method by showing the expected sensitivity in the response to the PC (21)(22)(23). Associated and appropriate measures of variability between tissue replicates should be defined (*e.g.,* if standard deviations are used they should be within the 95% viability confidence interval; SD ≤ 18% for the VRM).

INTERPRETATION OF RESULTS AND PREDICTION MODEL

27. The OD values obtained with each test sample can be used to calculate the percentage of viability normalised to NC, which is set to 100%. The cut-off value of percentage cell viability distinguishing irritant from non-classified test chemicals and the statistical procedure(s) used to evaluate the results and identify irritant chemicals, should be clearly defined, documented, and proven to be appropriate. The cut-off values for the prediction of irritation are given below:

> The test chemical is considered to be irritant to skin in accordance with GHS category 2 if the tissue viability after exposure and post-treatment incubation is less than or equal (≤) to 50%.
>
> Depending on country/regional regulatory requirements, the test chemical may be considered as "no category" if the tissue viability after exposure and post-treatment incubation is more than (>) 50%.

DATA AND REPORTING

Data

28. For each trial, data from individual replicate tissues (*e.g.,* OD values and calculated percentage cell viability data for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition means ± SD for each trial should be reported. Observed interactions with MTT reagent and coloured test chemicals should be reported for each

tested chemical.

Test Report

29. The test report should include the following information:

Test and Control Chemicals:

- Chemical name(s) such as CAS name and number, if known;
- Purity and composition of the chemical (in percentage(s) by weight);
- Physical-chemical properties relevant to the conduct of the study (*e.g.,* physical state, stability and volatility, pH, water solubility if known)
- Treatment of the test/control chemicals prior to testing, if applicable (*e.g.,* warming, grinding);
- Storage conditions,

Justification of the RhE model and SOP used.

Test Conditions

- Cell system used;
- Calibration information for measuring device, and bandpass used for measuring cell viability (*e.g.,* spectrophotometer);
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to:
  - i) Viability
  - ii) Barrier function
  - iii) Morphology
  - iv) Reproducibility and predictivity
  - v) Quality controls (QC) of the model
- Details of the test procedure used;
- Test doses used, duration of exposure and post treatment incubation period;
- Description of any modifications of the test procedure;
- Reference to historical data of the model. This should include, but is not limited to:
  - i) acceptability of the QC data with reference to historical batch data
  - ii) acceptability of the positive and negative control values with reference to positive and negative control means and ranges.
- Description of evaluation criteria used including the justification for the selection of the cut-off point(s) for the prediction model

Results:

- Tabulation of data from individual test samples;
- Description of other effects observed.

Discussion of the results.

Conclusion.

LITERATURE

1) United Nations (UN) (2007). Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second revised edition, UN New York and Geneva, 2007. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html]
2) OECD (2004). Test Guideline 404. OECD Guideline for the Testing of Chemicals. Acute Dermal Irritation/Corrosion. Updated Guideline, adopted April 24, 2002. Available at: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]
3) OECD (2004). Test Guideline 430. OECD Guideline for the Testing of Chemicals. *In Vitro* Skin Corrosion: Transcutaneous Electrical Resistance (TER). Adopted April 13, 2004. Available at: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]
4) OECD (2004). Test Guideline 431. OECD Guideline for the Testing of Chemicals. *In Vitro* Skin Model. Adopted April 13, 2004. Available at: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]
5) OECD (2006). Test Guideline 435. OECD Guideline for the Testing of Chemicals. *In Vitro* Membrane Barrier Test Method. Adopted July 19, 2006. Available at: [http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html]
6) ECVAM (2007). Performance Standards for applying human skin models to *in vitro* skin irritation. Available under "Download study document", at: [http://ecvam.jrc.ec.europa.eu]
7) OECD (2005). OECD Series on Testing and Assessment No. 34. Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Available at: [http://www.oecd.org/document/30/0,3343,en_2649_34377_1916638_1_1_1_1,00.html]
8) Fentem, J. H., Briggs, D., Chesné, C., Elliot, G. R., Harbell, J. W., Heylings, J. R., Portes, P., Roguet, R., van de Sandt, J.J.M. and Botham, P. (2001). A prevalidation study on *in vitro* tests for acute skin irritation. Results and evaluation by the Management Team. *Toxicol. In Vitro* 15, 57-93.
9) Portes, P., Grandidier, M. H., Cohen, C. and Roguet, R. (2002). Refinement of the EPISKIN protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. *Toxicol. In Vitro* 16, 765–770.
10) Kandárová, H., Liebsch, M., Genschow, E., Gerner, I., Traue, D., Slawik, B. and Spielmann, H. (2004). Optimisation of the EpiDerm test protocol for the upcoming ECVAM validation study on *in vitro* skin irritation tests. *ALTEX* 21, 107–114.
11) Kandárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D. and Spielmann H. (2005). The EpiDerm Test Protocol for the Upcoming ECVAM Validation Study on *In Vitro* Skin Irritation Tests – An Assessment of the Performance of the Optimised Test. *ATLA* 33, 351-367.
12) Cotovio, J., Grandidier, M.- H., Portes, P., Roguet, R. and G. Rubinsteen. (2005). The *In Vitro* Acute Skin Irritation of Chemicals: Optimisation of the EPISKIN Prediction Model within the Framework of the ECVAM Validation Process. *ATLA* 33, 329-349.
13) Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R. D., Elliot, G. R., Fentem, J. H., Heylings, J. R., Liebsch, M., Medina, J., Roguet, R.,van De Sandt, J. J. M., Wiemann, C. and Worth, A. (2002). Follow-up to the ECVAM prevalidation study on *in vitro* tests for acute skin irritation. ECVAM Skin Irritation Task Force Report 2. *ATLA* 30,109-129.
14) Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotović, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I. and Zuang, V. (2007). The ECVAM International Validation Study on *In Vitro* Tests for Acute Skin Irritation: Report on the Validity of the EPISKIN and EpiDerm Assays and on the Skin Integrity Function Test. *ATLA* 35, 559-601.

15) Hoffmann, S. (2006). ECVAM Skin Irritation Validation Study Phase II: Analysis of the Primary Endpoint MTT and the Secondary Endpoint. Available under "Downloads study documents", at: [http://ecvam.jrc.ec.europa.eu/index.htm accessed 10.12.2007]

16) Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang.V (2007) ECVAM International Validation Study on *In Vitro* Tests for Acute Skin Irritation: Selection of Test Chemicals. *ATLA* 35, 603-619.

17) Cotovio, J., Grandidier, M.-H., Lelièvre, D., Roguet, R., Tinois-Tessonneaud, E. and Leclaire, J. (2007). *In vitro* acute skin irritancy of chemicals using the validated EPISKIN model in a tiered strategy-Results and performances with 184 cosmetic ingredients, *AATEX*, Special Issue-proceedings from WC6. Vol.14, 351-358.

18) ECVAM (2008). Statement on the scientific validity of *in vitro* tests for skin irritation testing, issued by the ECVAM Scientific Advisory Committee (ESAC29), 5 November 2008. Available at: [http://ecvam.jrc.ec.europa.eu]

19) EC (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union L396/1 of 30.12.2006. OPOCE, Luxembourg.

20) Mosman, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immuno. Meth*. 65, 55-63.

21) EpiSkin™ SOP (2005). Version 1.6. Validation of the EpiSkin Skin Irritation Test - 42 Hours Assay For The Prediction of Acute Skin Irritation of Substances. Available under Download Study Documents, at: [http://ecvam.jrc.ec.europa.eu/]

22) EpiDerm™ SOP (2004). Version 7.0 Draft Standard Operating Procedure. *In Vitro* Skin Irritation Test: Human Skin Model. Model: EpiDerm™- 200. Available under Download Study Documents, at: [http://ecvam.jrc.ec.europa.eu]

23) SkinEthic RHE™ SOP (Draft). Will be available under Download Study Documents, at: [http://ecvam.jrc.ec.europa.eu]

24) Harvell, J. D., Lamminstausta, K, and Maibach H. I. (1995). Irritant contact dermatitis. In: Guin J. D. (ed.), Practical Contact Dermatitis, Mc Graw-Hill New York, pp 7-18.

25) ECVAM (2009). Statement on the "Performance Under UN GHS of Three *In Vitro* Assays for Skin Irritation Testing; and; the Adaptation of the Reference Chemicals and Defined Accuracy Values of the ECVAM Skin Irritation Performance Standards", issued by the ECVAM Scientific Advisory Committee (ESAC30), 25 March 2009. Available at: [http://ecvam.jrc.ec.europa.eu]

26) Griesinger C., Barroso J. and Zuang V. (2008). ECVAM background document on the recent adaptations of the ECVAM performance standards for *in vitro* skin irritation testing in the context of the drafting process of an EU test method and an OECD draft test guideline. Ispra, November 13, 2008.

27) Basketter, D. A., York, M., McFadden, J. P. and Robinson M. K. (2004). Determination of skin irritation potential in the human 4-h patch test. *Cont. Derm*. 51, 1-4.

28) Jirova, D., Liebsch, M., Basketter, D., Spiller, E., Kejlova, K., Bendova, H., Marriott, M. and Kandarova, H. (2007). Comparison of human skin irritation and photo-irritation patch test data with cellular *in vitro* assays and animal *in vivo* data. *AATEX*, Special Issue-proceedings from WC6. Vol. 14, 359-365.

ANNEX 1

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method.

Batch control chemical: Benchmark chemical producing a mid-range cell viability response of the tissue.

Cell viability: Parameter measuring total activity of a cell population *e.g.,* as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT ([3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue;), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of chemicals being examined.

ET_{50}: Can be estimated by determination of the exposure time required to reduce cell viability by 50% upon application of the marker chemical at a specified, fixed concentration, see also IC_{50}.

Infinite dose: Amount of test chemical applied to the *epidermis* exceeding the amount required to completely and uniformly cover the *epidermis* surface.

GHS (Globally Harmonized System of Classification and Labelling of Chemicals by the United Nations (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (1).

IC_{50}: Can be estimated by determination of the concentration at which a marker chemical reduces the viability of the tissues by 50% (IC_{50}) after a fixed exposure time, see also ET_{50}.

Me-too test: A colloquial expression for a test method that is structurally and functionally similar to a validated and accepted reference test method. Such a test method would be a candidate for catch-up validation. Interchangeably used with similar test method.

Performance Standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method.

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility.

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to t the accepted test, for all possible testing situations and chemicals.

Screening test: Often a rapid, simple test method conducted for the purpose of classifying chemicals into a general category of hazard. The results of a screening test generally are used for preliminary decision making in the context of a testing strategy (*i.e.,* to assess the need for additional and more definitive tests). Screening tests often have a truncated response range in that positive results may be considered adequate to determine if a chemical is in the highest category of a hazard classification system without the need for further testing, but are not usually adequate without additional information/tests to make decisions pertaining to lower levels of hazard or safety of the chemical

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method.

Skin irritation: The production of reversible damage to the skin following the application of a test chemical for up to 4 hours. Skin irritation is a locally arising, non-immunogenic reaction, which appears shortly after stimulation (24). Its main characteristic is its reversible process involving inflammatory reactions and most of the clinical characteristic signs of irritation (erythema, oedema, itching and pain) related to an inflammatory process.

Tiered testing strategy: Testing which uses test methods in a sequential manner; the test methods selected in each succeeding level are determined by the results in the previous level of testing.

ANNEX 2

PERFORMANCE STANDARDS FOR ASSESSMENT OF PROPOSED SIMILAR OR MODIFIED *IN VITRO* RECONSTRUCTED HUMAN *EPIDERMIS* (RhE) TEST METHODS FOR SKIN IRRITATION

INTRODUCTION

1. The purpose of Performance Standards (PS) is to communicate the basis by which new test methods, both proprietary (*i.e.,* copyrighted, trademarked, registered) and non-proprietary can be determined to have sufficient accuracy and reliability for specific testing purposes. These PS, based on validated and accepted test methods, can be used to evaluate the reliability and accuracy of other analogous test methods (colloquially referred to as "me-too" tests) that are based on similar scientific principles and measure or predict the same biological or toxic effect (7).

2. Prior to adoption of modified test methods, *i.e.,* proposed potential improvements to an approved test method, there should be an evaluation to determine the effect of the proposed changes on the test's performance and the extent to which such changes affect the information available for the other components of the validation process. Depending on the number and nature of the proposed changes, the generated data and supporting documentation for those changes, they should either be subjected to the same validation process as described for a new test, or, if appropriate, to a limited assessment of reliability and relevance using established PS (7).

3. Similar (me-too) or modified test methods proposed for use under this Test Guideline should be evaluated to determine their reliability and accuracy using chemicals representing the full range of the Draize irritancy scores. When evaluated using the 20 recommended Reference Chemicals of the PS (Table 1), the proposed similar or modified test methods should have reliability and accuracy values which are similar to those derived from the VRM (Table 2)(14). The reliability and accuracy values that should be achieved are provided in paragraphs 8- 9. Non-classified and classified (GHS category 2)(1) chemicals, representing relevant chemical classes are included, so that the reliability and accuracy (sensitivity, specificity and overall accuracy) of the proposed test method can be compared to that of the VRM. The reliability of the test method, as well as its ability to correctly identify GHS category 2 irritant chemicals and, depending on country/regional regulatory requirements, also its ability to correctly identify GHS no category chemicals (for member states that do not adopt optional category 3), should be determined prior to its use for testing new chemicals.

4. These PS are based on the ECVAM PS, updated according to the GHS on classification and labelling. The original PS defined for the VRM after the completion of the validation study (6) were based on the EU classification system. Due to the adoption of the GHS system for classification and labelling, which took place between the finalisation of the validation study and the completion of this Test Guideline, the PS have been updated. This update concerns mainly changes; (i), in the set of Reference Chemicals; and (ii), the defined accuracy values (24)(25).

PERFORMANCE STANDARDS FOR *IN VITRO* RhE TEST METHODS FOR SKIN IRRITATION

5. The PS comprise the following three elements (7):

I) Essential Test Method Components

II) Minimum List of Reference Chemicals

III) Defined Reliability and Accuracy Values

I) Essential Test Method Components

6. These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar or modified test method. These components include unique characteristics of the test method, critical procedural details, and quality control measures. Adherence to essential test method components will help to assure that a similar or modified proposed test method is based on the same concepts as the corresponding validated test method (7). The essential test method components are described in detail in paragraphs 16 to 21 of the Test Guideline and testing should be performed according to the following:

The general conditions (paragraph 15)

The functional conditions, which include:

viability (paragraph 16);
barrier function (paragraph 17);
morphology (paragraph 18);
reproducibility (paragraph 19); and,
quality control (paragraph 20)

II) Minimum List of Reference Chemicals

7. Reference Chemicals are used to determine if the reliability and accuracy of a proposed mechanistically and functionally similar or modified test method, proven to be structurally and functionally sufficiently similar to the VRM, or representing a minor modification of one of the three validated test methods, shows similar or better performance to that of the VRM (14). The 20 recommended Reference Chemicals listed in Table 1 include chemicals representing different chemical classes of interest, and are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant). The chemicals included in this list comprise 10 GHS category 2 chemicals and 10 non-categorised chemicals, of which 3 are optional GHS category 3 chemicals. Under this Test Guideline, the optional category 3 is considered as no category. The chemicals listed in Table 1 provide a representative distribution of the 58 chemicals used in the validation study of the VRM with regard to chemical functionality and physical state (16). These Reference Chemicals represent the minimum number of chemicals that should be used to evaluate the accuracy and reliability of a proposed similar or modified test method. In situations where a listed chemical is unavailable, other chemicals for which adequate *in vivo* reference data are available could be used. If desired, additional chemicals representing other chemical classes and for which adequate *in vivo* reference data are available may be added to the minimum list of Reference Chemicals to further evaluate the accuracy of the proposed test method.

Table 1. Reference Chemicals for determination of Accuracy and Reliability Values for Similar or Modified RhE skin irritation Test Methods[1]

Chemical	CAS RN	Physical state	*In vivo* score	VRM *in vitro* Cat.	GHS *in vivo* Cat.
1-bromo-4-chlorobutane	6940-78-9	Liquid	0	Cat. 2	No Cat.
diethyl phthalate	84-66-2	Liquid	0	No Cat.	No Cat.
naphthalene acetic acid	86-87-3	Solid	0	No Cat.	No Cat.
allyl phenoxy-acetate	7493-74-5	Liquid	0.3	No Cat.	No Cat.
isopropanol	67-63-0	Liquid	0.3	No Cat.	No Cat.
4-methyl-thio-benzaldehyde	3446-89-7	Liquid	1	Cat. 2	No Cat.
methyl stearate	112-61-8	Solid	1	No Cat.	No Cat.
heptyl butyrate	5870-93-9	Liquid	1.7	No Cat.	No Cat. (*Optional Cat. 3*)
hexyl salicylate	6259-76-3	Liquid	2	No Cat.	No Cat. (*Optional Cat. 3*)
cinnamaldehyde	104-55-2	Liquid	2	Cat. 2	No Cat. (*Optional Cat. 3*)
1-decanol*	112-30-1	Liquid	2.3	Cat. 2	Cat. 2
cyclamen aldehyde	103-95-7	Liquid	2.3	Cat. 2	Cat. 2
1-bromohexane	111-25-1	Liquid	2.7	Cat. 2	Cat. 2
2-chloromethyl-3,5-dimethyl-4-methoxypyridine HCl	86604-75-3	Solid	2.7	Cat. 2	Cat. 2
di-n-propyl disulphide*	629-19-6	Liquid	3	No Cat.	Cat. 2
potassium hydroxide (5% aq.)	1310-58-3	Liquid	3	Cat. 2	Cat. 2
benzenethiol, 5-(1,1-dimethylethyl)-2-methyl	7340-90-1	Liquid	3.3	Cat. 2	Cat. 2
1-methyl-3-phenyl-1-piperazine	5271-27-2	Solid	3.3	Cat. 2	Cat. 2
heptanal	111-71-7	Liquid	4	Cat. 2	Cat. 2
1,1,1-trichloroethane	71-55-6	Liquid	4	Cat. 2	Cat. 2

[1] The chemical selection is based on the following criteria; (i), the chemicals are commercially available; (ii), they are representative of the full range of Draize irritancy scores (from non-irritant to strong irritant); (iii), they have a well-defined chemical structure; (iv), they are representative of the chemical functionalities used in the validation process; and (v), they are not associated with an extremely toxic profile (*e.g.* carcinogenic or toxic to the reproductive system) and they are not associated with prohibitive disposal costs.

* Chemicals that are irritant in the rabbit but for which there is reliable evidence that they are non-irritant in humans (26)(27).

III) Defined Reliability and Accuracy Values

8. The accuracy (sensitivity, specificity and overall accuracy) of the proposed similar or modified test method should be similar to that of the VRM, taking into consideration additional information relating to relevance in the species of interest (Table 2), *i.e.,* sensitivity should be equal or higher (≥) than 80% with only 1-decanol and di-n-propyl disulphide being allowed to be misclassified against the rabbit classification (based on the mode of the classifications obtained from all participating laboratories), specificity should be equal or higher (≥) than 70%, and accuracy should be equal or higher (≥) than 75% (24). Although the sensitivity of the VRM calculated for the 20 Reference Chemicals listed in Table 1 is equal to 90%, the defined minimum sensitivity value required for any similar or modified test method to be considered valid is set at 80% since both 1-decanol (a borderline chemical) and di-n-propyl disulphide (a false negative of the VRM) are known to be non-irritant in humans (26)(27), although being identified as irritants in the rabbit test. However, of the 10 category 2 Reference Chemicals listed in Table 1, only 1-decanol and di-n-propyl disulphide are allowed to be misclassified by the proposed similar or modified test method against the rabbit test classification, based on the mode of the classifications obtained from all participating laboratories.

Table 2. Required predictive values for any similar or modified test method to be considered valid.

Sensitivity	Specificity	Overall Accuracy
80%	70%	75%

9. The calculation of the accuracy should be done on the basis of individual laboratory predictions, using all classifications obtained for the 20 Reference Chemicals in the different participating laboratories. The classification for each chemical in each laboratory should be obtained by using the mean value of viability over the different runs performed (minimum three valid runs).

Within-laboratory reproducibility

10. An assessment of within-laboratory variability should show a concordance of classifications (GHS category 2/no category) obtained in different, independent test runs of the 20 Reference Chemicals within one single laboratory equal or higher (≥) than 90%.

Between-laboratory reproducibility

11. An assessment of between-laboratory reproducibility is not essential if the proposed test method is to be used in one laboratory only. For methods to be transferred between laboratories, the concordance of classifications (GHS category 2/no category) obtained in different, independent test runs of the 20 Reference Chemicals between preferentially a minimum of three laboratories should be equal or higher (≥) than 80%.